

A Fast 3-D Object Recognition Algorithm for the Vision System of a Special-Purpose Dexterous Manipulator

Stephen H.Y. Hung

National Research Council of Canada
Ottawa, Ontario, Canada K1A 0R6

Abstract

A fast 3-D object recognition algorithm that can be used as a quick-look subsystem to the vision system for the SPDM is described. Global features that can be easily computed from range data are used to characterize the images of a viewer-centered model of an object. This algorithm will speed up the processing by eliminating the low level processing whenever possible. It may identify the object, or reject a set of bad data in the early stage, or create a better environment for a more powerful algorithm to carry the work further.

1. Introduction

As Canada's contribution to the International Space Station, the Mobile Servicing System (MSS) is being developed to support a number of major functions on the Station, including Station assembly and maintenance, attached payload servicing, transportation, payload handling, and astronaut extra-vehicular activity. Many of the functions to be performed by the MSS require the execution of complex dexterous manipulation operations beyond the capability of the Space Station Remote Manipulator System (SSRMS). This functional capability is provided by the Special Purpose Dexterous Manipulator (SPDM). The SPDM consists of a base, body, and two manipulator arms. The manipulator arms are of the order of 1.5 to 2.0 metres long, have seven rotary joints each, and terminate with a tool changeout mechanism (Fig. 1). The SPDM is intended to be operated primarily from the end of the SSRMS. Control of the SPDM will be effected by the operator through one of the MSS work stations. Initially, the SPDM will operate in a teleoperated fashion with some autonomous capability. To reduce the crew time required for servicing and maintenance operations, increased degrees of autonomy will be incorporated in the SPDM operations as the system evolves. Detailed description of the SPDM can be found in Ref. 1.

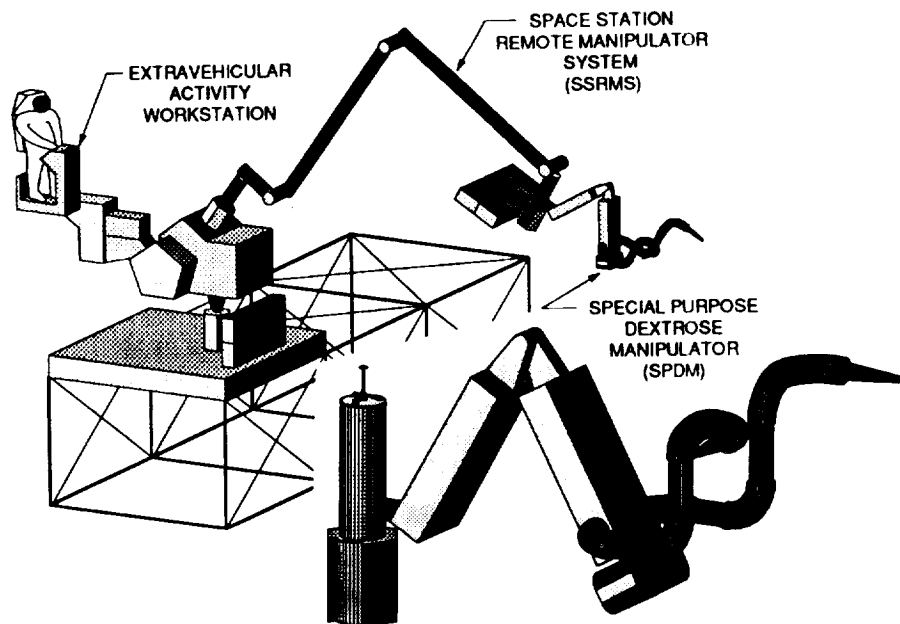


Figure 1.

The degrees of autonomy of the SPDM depend crucially on the vision capability. Vision sensors (video cameras, laser scanner) will be mounted near the manipulator arm ends. A vision system will also be mounted on the body to provide overall views of the work area of the manipulators. The vision systems will be able to identify objects (initially by reading labels), and to determine the position, the attitude, and rates of motion of a body relative to a known reference frame. Software and hardware provisions are being designed into the vision systems to enable enhancement of their capabilities to include more general pattern recognition and understanding of three-dimensional images. Three dimensional image data will be used in object identification and to update world models of objects for collision detection and avoidance. The data will be acquired with the stereo cameras or possibly a laser range finder. The algorithm presented in this paper is one of the possible methods designed for such applications.

2. Some Considerations

2.1. The space environment is characterized by very harsh lighting conditions, pronounced shadows, peculiar behaviour of light scattering, and the absorption bands in the near-infrared region of the earth background. All these are difficult for vision systems to handle. Thus, space vision systems should provide a mode independent of sunlight and be earth blind. Our experience suggests that the simplest approach would be to use laser range finders to collect range data for the space vision systems. Within the reach limit of the SPDM, laser range finders will work very well.

A laser range finder [2] has been under development in our laboratory for some years and has been used to collect a large quantity of range data [3]. It can capture a scene of 256 by 256 pixels within a second. Another compact wrist-mounted laser scanner [4] has also been successfully developed in our laboratory and has been mounted on the wrist of a PUMA robot arm to perform parts acquisition [5,6]. This compact scanner provides a single raster scan of range data. The robot arm must move in steps of small intervals to obtain the data for an area. Further development will allow range data over an area to be obtained by sweeping the wrist without moving the arm. Both of these systems are quite suitable for the vision system of SPDM.

The data obtained by both scanners have been used as the input data for extensive research work carried out in our laboratory on 3-D object recognition [7-13]. The algorithm proposed in this paper is based on the combination of some of these previous results.

2.2. In most space scenarios, the objects being viewed are unlabelled objects from a library of known and precisely defined objects. In our algorithm, we assume this to be the case and attempt to capitalize on it to reduce the computational effort of the space vision systems. Thus, we are adopting a model-based approach.

Model-based object recognition requires matching features measured in an observed image with models of objects. When 3-D objects may appear at any position and orientation, 3-D object models must be used. Usually, objects are modelled using either one object-centered representation or many viewer-centered descriptions (one for each viewpoint) [14]. In 3-D object-centered representations, objects are modelled by volumetric or surface models. Three-dimensional multiview representations model objects by a finite set of viewer-centered descriptions [15,16]. The major advantage of the multiview viewer-centered model over the object-centered representation is that the features extracted from images can be directly matched with the features associated with each viewpoint of the multiview model set.

However, the traditional approaches of either object-centered or viewer-centered models in object recognition have a common bottleneck. Both of them must extract local features such as edges, corners, surfaces, and the relations between them as basic data to describe an object or a class of objects in the reference data set. The major drawbacks of this kind of approach are the difficulty and large computational effort required in extracting the local features and their relations, as well as the complicated processing of the comparison. The combination of these drawbacks usually makes such methods slow and difficult to implement in realtime and, in some cases, they lead to a combinatorial explosion.

The algorithms based on the traditional approach may be powerful, but should not be used unnecessarily. There are many cases in which the use of these methods is obviously not desirable or at least should be postponed. The following are some of those cases:

- A . If the observed object is obviously different from the object we are looking for, then the observed data should be rejected immediately without going through all the complicated low level processings.
- B . The data are obtained from a poor viewing angle, and there is not enough information to identify the object. Such data should be discarded before the low level processing is involved and the system should be directed to look from another angle.
- C . In many cases an object can be identified easily by looking at it from different viewpoints and relying only on some obvious global features without complicated low level processing.

This suggested that there should be a quick-look step before the low level processing to determine whether the full scale processing is actually needed. If the full processing is needed, then the quick-look step should be used to make sure that bad viewing angles have been avoided, and possible candidates have been narrowed down by eliminating those obviously unsuitable ones. This is true for all vision systems; it is especially true for the space vision systems. We should use a quick-look step to minimize effort and speed up the processing.

However, any such quick-look algorithm without low level processing to extract local features must rely only on the global features. The limitation of such approach is known. We must be able to separate the potential object either from the background or from the other objects.

2.3. The SPDM will be working autonomously in a space of about a cube of 2 metres in each dimension due to the reach limit of the arms. In such a space, the objects that SPDM will deal with can be roughly divided into three categories:

- A. The object is isolated or can be isolated easily, such as a part or a tool floating in the space.
- B. The object is attached to a big object, and must be separated from the background, such as the ORU.
- C. Two or more objects are stuck together, one may be over the other, and may or may not be separated physically.

The first two cases can be handled easily and should not be a problem. We must derive some kind of procedures to handle the case C. The most intuitive one is examination from a different angle, locating the top-most object, and slicing it off from the rest. In the space and the use of range data actually make such attempt easier.

In the following sections we will describe a quick-look algorithm and how to make it work. The main principle is that each step in the system is trying to create a better environment for the next step, until the object is identified.

3. Basic Assumptions and Brief Outline of our Approach

We assume that a laser range finder is mounted on the end of each arm and in the top of the lower body. The range finders can be programmed to scan a single profile or full resolution of 256 by 256 pixels. A reference plane can be set at any distance along the Z-axis. The readings of X, Y and Z are given in millimetres from the reference coordinate system at which the Z values of the reference plane are all zeros (Fig. 2).

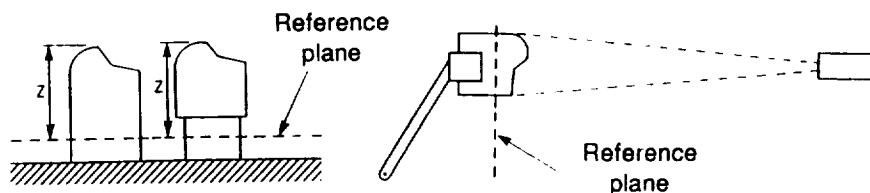


Figure 2.

The whole system can be divided into three subsystems:

3.1. The first stage is to locate the object and to determine the viewing angle and the reference plane. At this stage, we are using only the range finders in the single profile mode. They may need to scan several times to locate the object.

3.2. The second stage is a quick-look step. This step must be fast, not involving any low level processing, and able to be repeated easily without much delay to the whole processing. This subsystem should be able to eliminate the most obvious unsuitable candidates and determine if the input data can be identified, be rejected, or if this quick-look processing should be repeated by looking from another angle, or if the secondary subsystem should be invoked.

3.3. The secondary subsystem. When the quick-look subsystem fails to give a conclusive answer, a more powerful subsystem must be called in to carry the processing further. In this stage, low level processing will be involved, and features will be extract. Any traditional or otherwise method can be adopted as one see fit. The processing may be more complicated, but will only be invoked when necessary, and under the more favourable conditions prepared by the quick-look subsystem. This stage is beyond the scope of this paper and will not be discussed further.

4. The Algorithm for the Quick-look Subsystem

The algorithm for the quick-look step in this system is described in [12]. It is based on multiview viewer-centered representation. Each image of an object from a possible viewing angle is characterized by four global features which can be quantified and can be readily calculated from the raw data of the image. The features of images from all possible viewpoints of an object are stored as the model of the object. The collection of all models makes up the library of models for the quick-look subsystem.

The four features chosen are the approximate size and shape of the silhouette, the histogram of Z-values and the distribution of Z in the (x,y) plane.

4.1. The size of the silhouette is considered as the most obvious feature to distinguish an image of a particular viewing position of an object. Since the scanner is making a reading of Z at a constant interval along both the X and Y axes, the size of a silhouette is simply the number of points in an image. It is invariant under in-plane rotation and translation.

4.2. The shape of the silhouette is the next obvious feature to be considered. A moment invariant is chosen to represent the shape of the silhouette. It is the IV2D below:

$$\phi_{pq} = \sum_x \sum_y (x - \bar{x})^p \cdot (y - \bar{y})^q \quad (\text{central moment})$$

where \bar{x} and \bar{y} are the mean values of x and y, respectively

$$\psi_{pq} = \phi_{pq} / \phi_{00}^r, \quad \text{where } r = \frac{1}{2}(p + q) + 1 \quad (\text{normalized central moment})$$

$$\text{IV2D} = (\psi_{20} + \psi_{02}), \quad P_2 = \text{IV2D} \times 1000.0$$

Since the value of IV2D generally is too small, we multiply it by 1000.0 and then use it as a parameter P₂. Parameters P₂, P₃ and P₄ are independent of size, in-plane rotation and translation. P₂ is solely determined by the shape.

4.3. The histogram of Z-values is a good indicator of the volatility of the visible surfaces. The best parameter to describe the histogram is its standard deviation, i.e.,

$$\sigma = \sqrt{\frac{\sum_{x,y} (z - \bar{z})^2}{N}}, \quad z = f(x,y) \quad \text{where: } \bar{z} \text{ is the mean value of } z = f(x,y), \\ N \text{ is the number of points.}$$

The standard deviation is also independent of the height of the object. The standard deviations of two different objects can be compared without knowing the heights of the two objects.

4.4. The distribution of Z-values in the (x,y) plane is a good descriptor of the visible portion of a 3-D object at a certain viewing angle, i.e., a 3-D image. The quantity used to represent the distribution of Z is an invariant of moments indicated as IV3D:

$$\mu_{pq} = \sum_{x,y} (x - \bar{x})^p \cdot (y - \bar{y})^q \cdot f(x,y), \quad z = f(x,y) \quad (\text{central moment})$$

\bar{x} and \bar{y} are the mean values of x and y, respectively

$$\eta_{pq} = \mu_{pq} / \mu_{00}^\gamma, \quad \text{where } \gamma = \frac{1}{2}(p+q) + 1 \quad (\text{normalized central moment})$$

$$\text{IV3D} = (\eta_{20} + \eta_{02}), \quad P_4 = \text{IV3D} \times 1000.0$$

P₄ is defined as IV3D multiplied by 1000.0. P₄ is not independent of the Z-values. The comparison of P₄ of two images will be meaningful only if the two images are the same distance from the reference plane. The P₄ at a certain height can always be computed from the P₄ at a different height, provided that the average value of Z (AVZ) is also available. Thus, the P₄ can always be compared at the same height.

The four parameters are:

- P₁ = Size of the Silhouette (Number of Points)
- P₂ = Shape of the Silhouette (IV2D × 1000.0)
- P₃ = Volatility of the Visible Surface (σ, Standard Deviation of the Z-values)
- P₄ = Distribution of Z-values in (x,y) plane (IV3D × 1000.0)

For example, the four parameters of the grapple feature when scanned directly from the top centre are shown in Fig. 3. Because this set of data includes the round platform, this set of parameters does not really show the special character of this feature. Since the highest point in the data or the distance to the platform is known, the platform can be sliced off and a set of data obtained as in Fig. 4. The parameters computed from this set of data give a better representation of the grapple from this particular viewpoint.

One of the important reasons for choosing these four parameters is that all four parameters and the average value of Z (AVZ) can be readily computed by going through the data only once. Thus, low level processing is not needed. The first three parameters are used to classify the images of objects in a three-dimensional feature space. Once the four parameters of the input data are obtained, an allowable error limit is set for each of the first three parameters. The input data are characterized by a small cube in the feature space in such a way that the centre of the cube is determined by the first three parameters, and each dimension is the allowable error. The search is simply to determine if this cubic actually intersects any clusters (characterized images). If there is no intersection, then the input data are rejected. Otherwise, the possible candidates and the input data will be adjusted to have the same AVZ; then the P₄ will be compared.

If the result is inconclusive, then examination from another viewing angle will proceed. Usually the second or the third try will give a definite answer (identified, rejected, or secondary subsystem should be invoked). If after several (five or six) tries there is still no conclusive answer, then the secondary subsystem should be invoked.

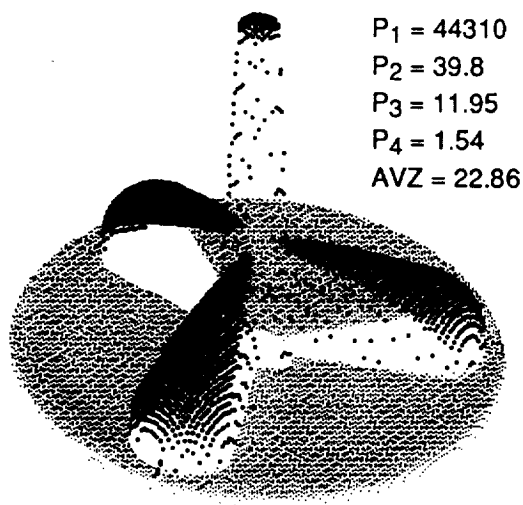


Figure 3.

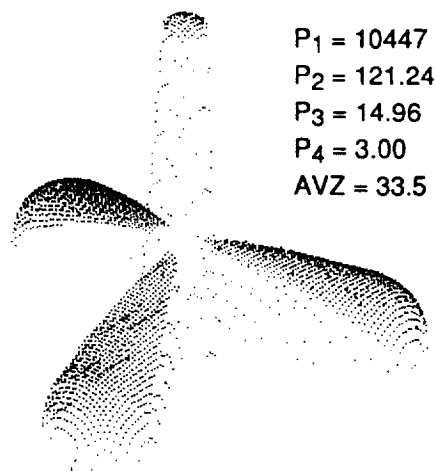


Figure 4.

5. The Searching Subsystem and the Secondary Subsystem

The major mandate of the searching subsystem is to locate the object. If there is something to look at, then the subsystem must determine the viewing angle, how to set the reference plane, and the approximate area to be scanned to cover the possible object.

The analysis of the single profile of the laser scanner can be used to determine if there is an object. If there is a possible object, then a few cross or parallel scans will be enough to decide the approximate area to cover the whole object. The viewing angle and the reference plane should then be set to separate the object from the background or other objects.

The rule of thumb is that whenever possible we should scan from an angle that is perpendicular to the major surface or the background, and the reference plane should be set to cross the jump edges in a single profile of range data, as in Fig. 5.

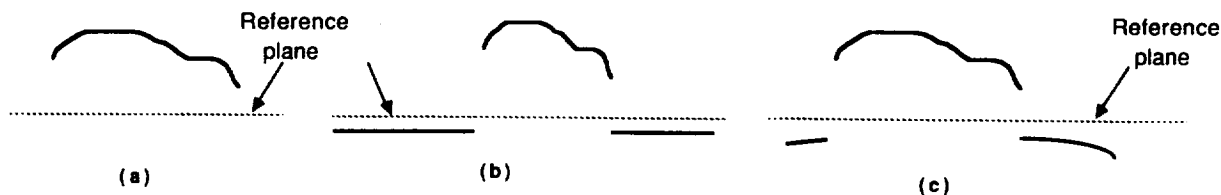


Figure 5.

When two or more objects stick together, the best we can do is to slice off the top-most object from the rest. In range data, the top-most object is usually separated with the rest of the scene by the jump edges. The object at the bottom will become the top-most one when looked at from a different angle, as in Fig. 6. If the two pieces are not stuck together physically, the bottom one may be identified by moving the top pieces away first. This stage should provide the quick-look subsystem with a set of workable data.

A separated portion in the model of each object is specially set aside for the searching subsystem. In this portion, major dimensions are listed and can be directly referred to without being extracted from the 3-D model.

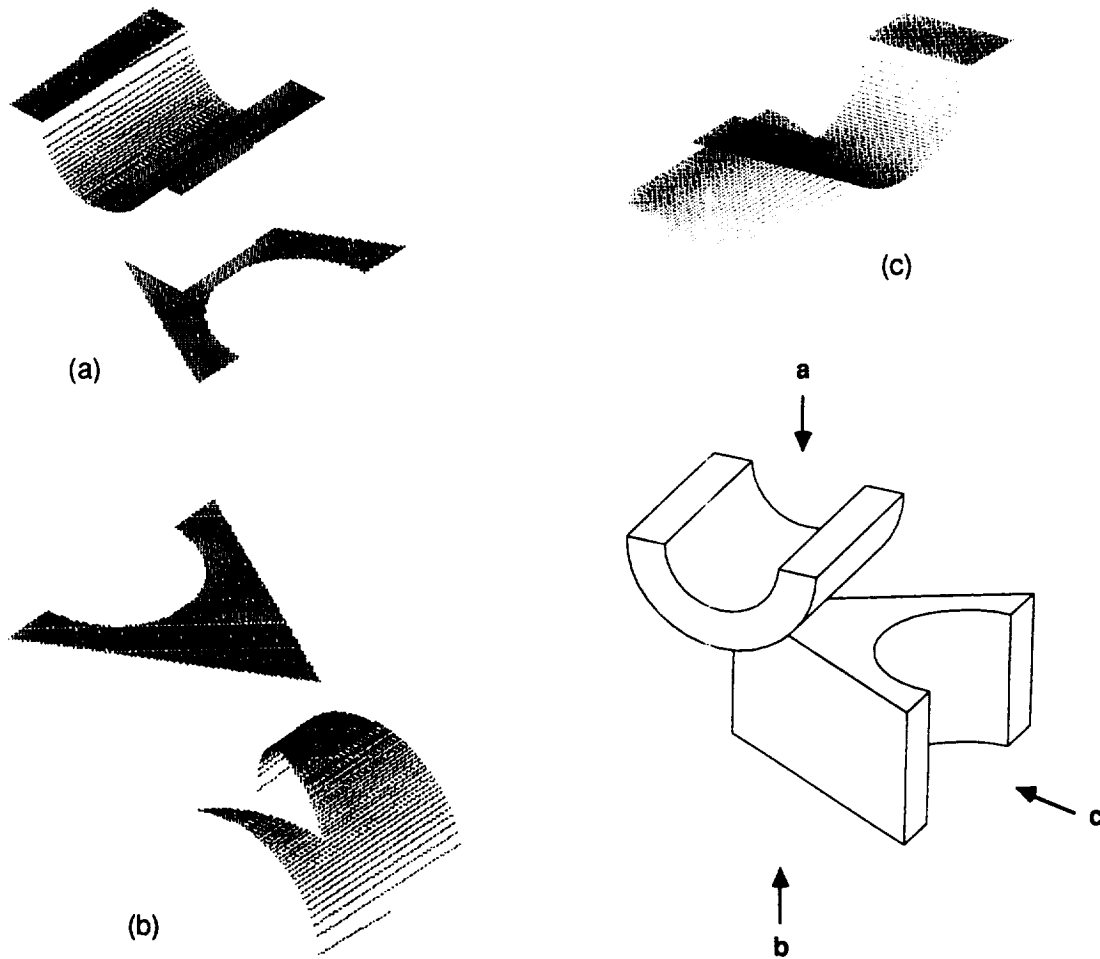


Figure 6.

A few points can be made about the secondary subsystem:

- A . In most cases, the secondary subsystem is called to narrow down several possible candidates to a single one. Since the observed object is known by now to possibly be identified with one object in a group of similar objects, a tailored algorithm may be enough to handle the situation. Extracting of some special features from the 3-D model will provide everything needed to do the job. Thus, a CAD-based representation may be preferred to a volumetric representation, although the latter may make it easier to compare two complete objects.
- B . A multiview representation in this stage may be neither necessary nor desirable. The model for the quick-look subsystem is a multiview representation in a simpler form. The possible advantage of the multiview model is already taken. The ease of extraction of a special feature from a 3-D model is the most important concern at this stage, the object-centered model seems better in this respect.
- C . The most frequent reason for the failure of the quick-look subsystem is that the object has not been properly separated from the scene. Before the full scale of low level processing is involved, a limited low level processing which extracts only the jump edges can be applied. The result may help separate the top-most object from the rest of the scene, and conclusive result may be obtained by applying the quick-look steps again.

The model of an object for the system is shown in Fig. 7. Besides the information described previously, there is a place for some special instructions in each stage. At the beginning, the special instructions are put in by the system designer. In the future, the system will have a self-improving ability to update those instructions as the system learns from its experience. This may be in a distance future, but it is a must for a vision system to work properly.

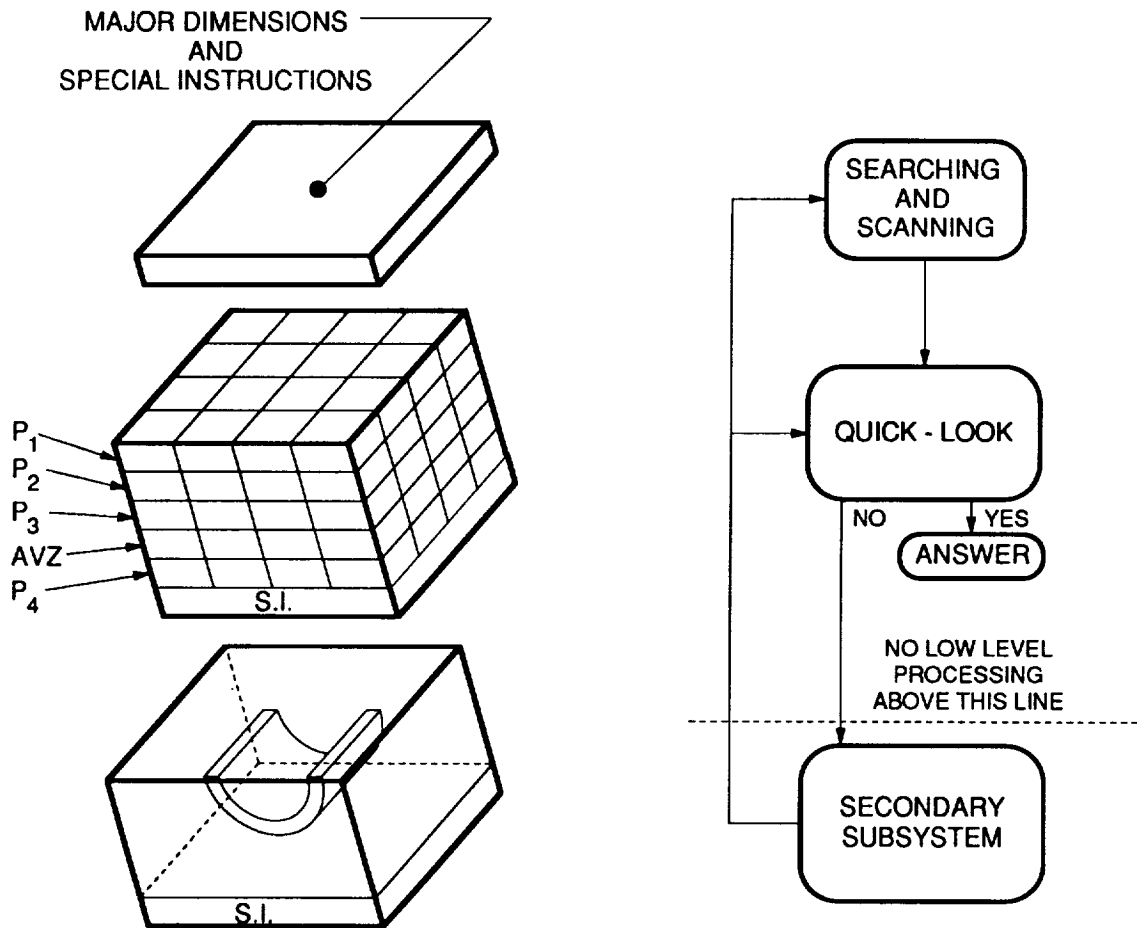


Figure 7.

6. Example

The procedure for locating and confirming the grapple feature is described as follows:

1. Assuming that the grapple feature is located in the end of a large cylindrical body. The SPDM is guided by the SSRMS to the vicinity of the large body. The searching subsystem must determine the orientation of the body and lead the SPDM to one of its end by analyzing some single profile data (Fig. 8).
2. When the SPDM is at one of the ends of the body, at least two more single profile data must be obtained to determine whether the SPDM is at the top centre and perpendicular to the platform of the end of the body. The reference plane is set to just beyond the platform.

3. After scanning is completed, the set of range data is read once and the four parameters are computed. If the SPDM is at the right end, then the parameters should be closed to the set in Fig. 3, provided the viewing angle is not far off the top centre (within few degrees).
4. This situation should be further confirmed by slicing off the platform and computing the parameters for the remaining data. The parameters should be similar to the set in Fig. 4. This will also confirm that the SPDM is viewing from the top centre and perpendicular to the platform. Otherwise, the parameters computed from the remaining data (Fig. 9) after the slicing off action cannot match the set in Fig. 4.

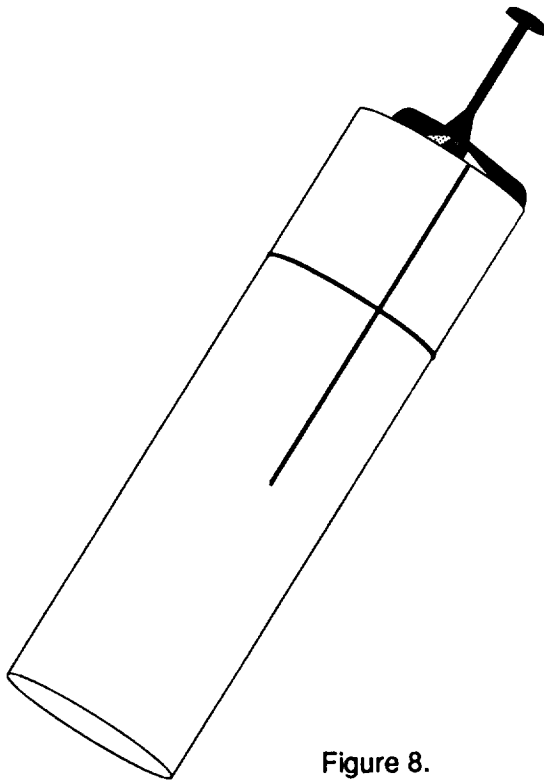


Figure 8.

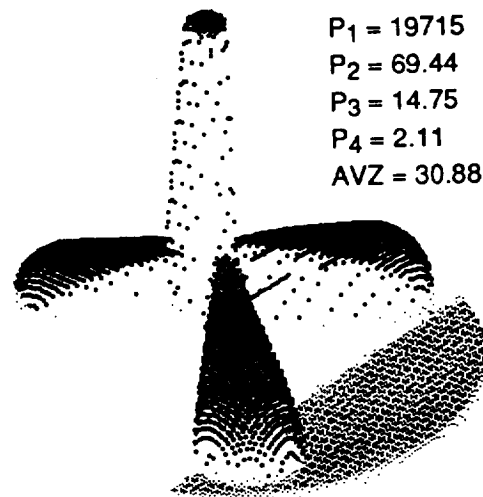


Figure 9.

7. Conclusion

A vision system has been proposed for the SPDM. Range data are chosen as the input data for this system. The emphasis is on a quick-look step that should be implemented before the low level processing becomes involved. By considering four aspects of the images of an objects in range data, the quick-look algorithm can speed up the processing by identifying the object, rejecting the data in the early stage, or reducing the number of possible candidates remaining in the field. A search step is applied to pave the way for this quick-look algorithm.

Major effort for implementing this quick-look algorithm is now being placed on the building of the library of models. This involves scanning the objects from many different angles to obtain the range data, computing the parameters, and then discarding the data and storing only the parameters. However, all these are done in off-line operation, little on-line computation will be required to obtain a conclusion when the input data are given.

This is only a preliminary study on the feasibility of a possible space vision system for the SPDM. Much must be done to make it practically worthwhile.

8. References

1. Hunter, D.G., "An Overview of the Space Station Special Purpose Dexterous Manipulator (SPDM)". National Research Council of Canada, NRCC No. 28817, April, 1988.
2. Rioux, M., "Laser Range Finder Based on Synchronized Scanners". Appl. Opt. 23(21), pp. 3837-3844; 1984. NRC 24529.
3. Rioux, M., and Cournoyer, L., "The NRCC Three-dimensional Image Data Files". CNRC 29077.
4. Rioux, M., Bechthold, G., Taylor, D. and Duggan, M., "Design of a Large Depth of View Three-dimensional Camera for Robot Vision". Optical Engineering, 26(12), pp. 1245-1250; 1987. NRC 29130.
5. Roth, G., and O'Hara, D.H., "A Holdsite Method for Parts Acquisition Using a Laser Rangefinder Mounted on a Robot Wrist". Proc. IEEE Int'l. Conf. on Robotics and Automation, Raleigh, North Carolina, March, 1987, pp. 1517-1523.
6. Archibald, C., "Processing Single Range Profiles from a Wrist Mounted Laser Range Finder". Proc. Vision Interface'88, Edmonton, Alberta, June, 1988, pp. 13-18.
7. Archibald, C., and Rioux, M., "Witness: A system for Object Recognition Using Range Images". NRC ERB-986, Jan. 1986.
8. Oka, R., Kasvand, T. and Rioux, M., "Recognition of Three-dimensional Objects Based on the Cross-angle Transform". NRC ERB-980, April 1986.
9. Dhome, M., and Kasvand, T., "Polyhedra Recognition by Hypothesis Accumulation". IEEE Trans. PAMI-9(3), pp. 429-438, 1987. NRC 29137.
10. Hospital, M., Yamada, H., Kasvand, T., and Umeyama, S., "3D Curve Based Matching Method Using Dynamic Programming". Proc. 1987 IEEE First International Conf. on Computer Vision, London, England. June 8-11, 1987. pp. 728-732.
11. Abdelmalek, N.N., "A Global System for matching 3-D Range Data Objects". Proc. Vision Interface'88, Edmonton, Alberta, Canada, June 6-10, 1988. pp. 129-134.
12. Hung, S.H.Y., "A Practical Approach of 3-D Object Recognition in Range Data". Proc. First Int'l. Conf. Industrial & Eng. Appl. of Artificial Intil. & Expert Systems. UTSI, Tulaoma, Tennessee, June 1-3, 1988. pp. 1131-1138.
13. Merritt, C., Archibald, C., and Ng, T., "Pose Determination of a Satellite Grapple Fixture Using a Wrist-mounted Laser Range Finder". Proc. SPIE 1002, Intelligent Robots and Computer Vision, Cambridge, MA. Nov. 1988.
14. Chin, R.T., and Dyer, C.R., "Model-Based Recognition in Robot Vision". Computing Surveys, Vol. 18, No. 1, pp 67-108. March 1986.
15. Ikeuchi, K., "Generating an Interpretation Tree from a CAD Model for 3D-Object Recognition in Bin-Picking Tasks". Int'l. J. Computer Vision, pp 145-165, 1987.
16. Kron, M.R., and Dyer, C.R., "3-D Multiview Object Representations for Model-based Object Recognition". J. Pattern Recognition, Vol. 20, No. 1, pp 91-103, 1987.